

Terabase Metagenome Sequencing of Grassland Soil Microbiomes



Nelson WC, Anderson LN, Wu R, McDermott JE, Bell SL, Jumpponen A, Fansler SJ, Tyrrell KJ, Farris Y, Hofmockel KS, Jansson JK. 2020. Terabase metagenome sequencing of grassland soil microbiomes. *Microbiol Resour Announc* 9: e00718-20. <https://doi.org/10.1128/MRA.00718-20>.

ABSTRACT

As part of the Pacific Northwest National Laboratory's (PNNL) Science Focus Area program we are investigating the impact of environmental change on microbial community function in grassland soils. To enable a comprehensive survey of the metabolic potential of complex soil microbiomes, we performed ultra-deep metagenome sequencing, collecting >1 terabase of sequence data from grassland soils representing different precipitation regimes. Frequent precipitation regime soil (rain-fed and tile-drained agriculture) was collected from Iowa State University's Comparison of Biofuel Systems (COBS).

METAGENOME BIOPROJECT ACCESSION

[IA-TmG.1.0](#)

FIELD SITE

Comparison of Biofuels Systems (COBS)
Boone County, IA, USA

Iowa State University Agronomy and Agricultural Engineering Research Farm, Comparison of Biofuel Systems (COBS), and experimental site located on the South Reynoldson Farm in Boone County, Iowa (est. 2008-present). COBS is a large-scaled bioenergy cropping system designed for producing high biomass/biofuel feedstock yields for maximal net energy return. The COBS site consists of a conventional corn-soybean cash grain system (C2/S2), a continuous corn grown for grain and stover removal with (CCW) and without (CC) a rye crop cover, and multi-species perennial crops grown for aboveground biomass with (PF) and without (P) fertilizer. This field site is considered a "no-till" system (harvested in late fall each year) and maintains an infrastructure for continuously monitored flow-volume measurements and flow-integrated sampling of subsurface drainage from each plot.

BIOSAMPLE RECORD

Identifiers: BioSample Name: IA-TmG
Assembly Name: IA-TmG.1.0
Pooled Lab IDs: IA-1, IA-2, IA-3, IA-4, IA-5, IA-6, IA-7, IA-8, IA-9

Organism: soil metagenome (unclassified sequence)

Package: [MIMS: metagenome/environmental_soil; version 5.0](#)

Attributes:

collection date	2017-11-07
depth	0-20 cm
elevation	317 m
broad-scale environmental context	ENVO:01000177 (grassland biome)
local-scale environmental context	ENVO:00005750 (grassland soil)
environmental medium	ENVO:00001998 (soil)
geographic location	GAZ:00002459:GAZ:00004438:GAZ:22225729
latitude and longitude	41.92 N 93.75 W
annual and seasonal precipitation	934 cm
annual and seasonal temperature	8.9 °C; 15 °C (high), 3 °C (low)
pH:	6.7
organic matter content:	1.724 (51 g kg ⁻¹)
Soil Type:	Webster series
Soil Map:	Mollisol (aquolls)
Soil Class:	Typic Endoaquolls (fine-loamy, mixed, superactive, mesic)

Description: Complete Terabase metagenome of grassland soil microbiome collections from COBS field site in Boone County, IA.

BioProject: PNNL Soil Microbiome SFA 'Phenotypic Response of the Soil Microbiome to Environmental Perturbations' (70880)

METAGENOMIC STRATEGY

DNA Isolation Method: [DNeasy PowerSoil Kit](#)

Sequencing Platform: Illumina HiSeq X (pyrosequencing)

Assembly Method: MetaHipMer (version 1.0-629-g3bce979f-20190325)

Assembly Platform: <https://docs.nersc.gov/systems/cori/>

FINAL ASSEMBLY STATISTICS

Minimum Scaffold Length	Number of Scaffolds	Number of Contigs	Total Scaffold Length	Total Contig Length	Contig Coverage
All	77,470,427	77,495,255	29,227,259,201	29,226,305,617	100.00%
100	77,470,427	77,495,255	29,227,259,201	29,226,305,617	100.00%
250	48,609,201	48,631,527	22,848,466,117	22,847,593,257	100.00%
500	12,077,304	12,081,050	10,676,414,657	10,676,265,906	100.00%
1 KB	2,599,812	2,600,921	4,393,212,492	4,393,167,054	100.00%
2.5 KB	289,845	290,028	1,152,748,070	1,152,740,230	100.00%
5 KB	47,664	47,701	359,982,354	359,980,614	100.00%
10 KB	6,379	6,382	91,875,307	91,875,164	100.00%
25 KB	308	309	10,819,282	10,819,258	100.00%
50 KB	30	31	1,897,726	1,897,702	100.00%

DATA PACKAGE CONTENTS

MIMS Environmental/Metagenome Standard Information (metadata)

- [IA-TmG.1.0_MIMS.me.soil.5.0 \(.xlsx\)](#)

Environmental metagenome sequence (unclassified sequences) package include attributes defined by the Genome Standards Consortium (GSC) to formally describe and standardize sample metadata.

COBS_IA-TmG.1.0_PlotMap.pdf

Field site plot location blocks and GPS coordinate information for collected soil cores.

Raw Metagenomic Data

- [RawReads_IA-TmG.1.0 \(704.44 GB; 18 items\)](#)

Name format:

<field site>-<sequencing lane>_<paired ends (forward and reverse) 1/2>.fastq.gz

IA-1_R1.fastq.gz
IA-1_R2.fastq.gz
IA-2_R1.fastq.gz
IA-2_R2.fastq.gz
IA-3_R1.fastq.gz
IA-3_R2.fastq.gz
IA-4_R1.fastq.gz
IA-4_R2.fastq.gz
IA-5_R1.fastq.gz
IA-5_R2.fastq.gz
IA-6_R1.fastq.gz
IA-6_R2.fastq.gz
IA-7_R1.fastq.gz

IA-7_R2.fastq.gz
IA-8_R1.fastq.gz
IA-8_R2.fastq.gz
IA-9_R1.fastq.gz
IA-9_R2.fastq.gz

Metagenomic Assembly Files

- [MetaHipMer_Assembly_IA-TmG.1.0 \(31.08 GB; 12 items\)](#)

File format:

.cov (average coverage across scaffolds)

IA-1.cov
IA-2.cov
IA-3.cov
IA-4.cov
IA-5.cov
IA-6.cov
IA-7.cov
IA-8.cov
IA-9.cov
IA-TmG_scaffold.fasta (assembled contigs)
IA.assembly.README (assembly parameters)

CDS and Functional Annotation Files

- [FunctionalAnnotation_IA-TmG.1.0 \(749.5 MB; 2 items\)](#)
IA-TmG.scaffold_2500.gff (annotation of biological sequences)
IA-TmG.scaffold_2500.prodigal.faa (predicted protein sequences)

DATA TAGS

Terabase metagenome, metaphenome, soil microbiome, SFA, Iowa, IA, COBS, soils, ecosystem, irrigated agriculture, marginal land, moisture regime, grasslands, environmental perturbation, intermediate moisture, rain-fed, microbial community, earth systems science, biology, data package, metadata, Illumina HiSeq X, MetaHipMer, BBTtools, Prodigal, NERSC, cori platform, raw reads, assembly, protein sequence predictions, EggNOG, MIMS.soil.5.0

Acknowledgements

This research was supported by the Department of Energy Office of Biological and Environmental Research (BER) and is a contribution of the Scientific Focus Area “Phenotypic response of the soil microbiome to environmental perturbations” (70880). and the EMSL/JGI FICUS award ‘Phenotypic Response of the Soil Microbiome to Environmental Perturbations’ (51227). PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

Address Correspondence

[Janet K. Jansson](#)

Biological Systems Science
Lab Fellow, Biologist
Pacific Northwest National Laboratory
PO Box 999
MSIN: J4-18
Richland, WA 99352
janet.jansson@pnnl.gov